

# Kinetic analysis of the role of the tyrosine 13, phenylalanine 56 and glutamine 54 network in the U1A/U1 hairpin II interaction

Michael J. Law, Eric J. Chambers<sup>1,2</sup>, Phinikoula S. Katsamba, Ian S. Haworth<sup>1,2</sup> and Ite A. Laird-Offringa\*

Departments of Biochemistry and Molecular Biology and of Surgery and <sup>1</sup>Department of Pharmaceutical Sciences, School of Pharmacy and <sup>2</sup>Department of Biochemistry and Molecular Biology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089-9176, USA

Received November 3, 2004; Revised April 15, 2005; Accepted May 1, 2005

## ABSTRACT

**The A protein of the U1 small nuclear ribonucleoprotein particle, interacting with its stem-loop RNA target (U1hplI), is frequently used as a paradigm for RNA binding by recognition motif domains (RRMs). U1A/U1hplI complex formation has been proposed to consist of at least two steps: electrostatically mediated alignment of both molecules followed by locking into place, based on the establishment of close-range interactions. The sequence of events between alignment and locking remains obscure. Here we examine the roles of three critical residues, Tyr13, Phe56 and Gln54, in complex formation and stability using Biacore. Our mutational and kinetic data suggest that Tyr13 plays a more important role than Phe56 in complex formation. Mutational analysis of Gln54, combined with molecular dynamics studies, points to Arg52 as another key residue in association. Based on our data and previous structural and modeling studies, we propose that electrostatic alignment of the molecules is followed by hydrogen bond formation between the RNA and Arg52, and the sequential establishment of interactions with loop bases (including Tyr13). A quadruple stack, sandwiching two bases between Phe56 and Asp92, would occur last and coincide with the rearrangement of a C-terminal helix that partially occludes the RRM surface in the free protein.**

## INTRODUCTION

The RNA recognition motif (RRM) is the third most common domain in human proteins, based on the sequence of the

human genome (1). The  $\beta$ - $\alpha$ - $\beta$ - $\beta$ - $\alpha$ - $\beta$  secondary structure of this domain assembles into an RNA-binding platform, consisting of a four-stranded anti-parallel  $\beta$ -sheet supported by two  $\alpha$ -helices [reviewed in (2–5)]. The most highly conserved regions of the RRM domain are two tracts of 8 and 6 amino acids, respectively, dubbed ribonucleoprotein consensus sequences 1 and 2 (RNP-1 and RNP-2, Figure 1A). The pervasiveness of RRM domains is probably due to their versatility; their presence in one to four copies in hundreds of RNA-binding proteins allows binding to a plethora of RNAs exhibiting a wide variety of sequences and structures. RNA binding mediated by RRMs can occur with very different affinities, reflecting the varied roles that RRM domain proteins play in the cell. These roles run the gamut from transiently RNA-associated chaperonin proteins such as nucleolin (6) to very stably RNA-associated building blocks of RNA-protein machinery, such as U1A in the U1 small nuclear ribonucleoprotein particle (U1 snRNP) (7). The participation of RRM-containing proteins in RNA-based gene regulation at all levels, as well as their role as building blocks of vital pieces of cellular machinery, underscores the importance of understanding the mechanism by which these domains mediate RNA binding. The elucidation of this mechanism is the focus of this study.

Crucial to the ability of RRM proteins to bind tightly to RNA are conserved aromatic residues that occur in RNP-1 and RNP-2 in most RRM domains and that lie centrally in the RNA-binding platform (3–5,8). Four aromatic residues are generally present in the consensus tracts—one in RNP-2 and three in RNP-1 (Figure 1A). In U1A, one of the three conserved RNP-1 residues (position 54) is a non-aromatic residue: a glutamine. Although their presence in most RRM domains suggests a critical role of the four aromatic residues in RNA binding, only two of the four residues appear to contact RNA targets, based on the solved structures of numerous

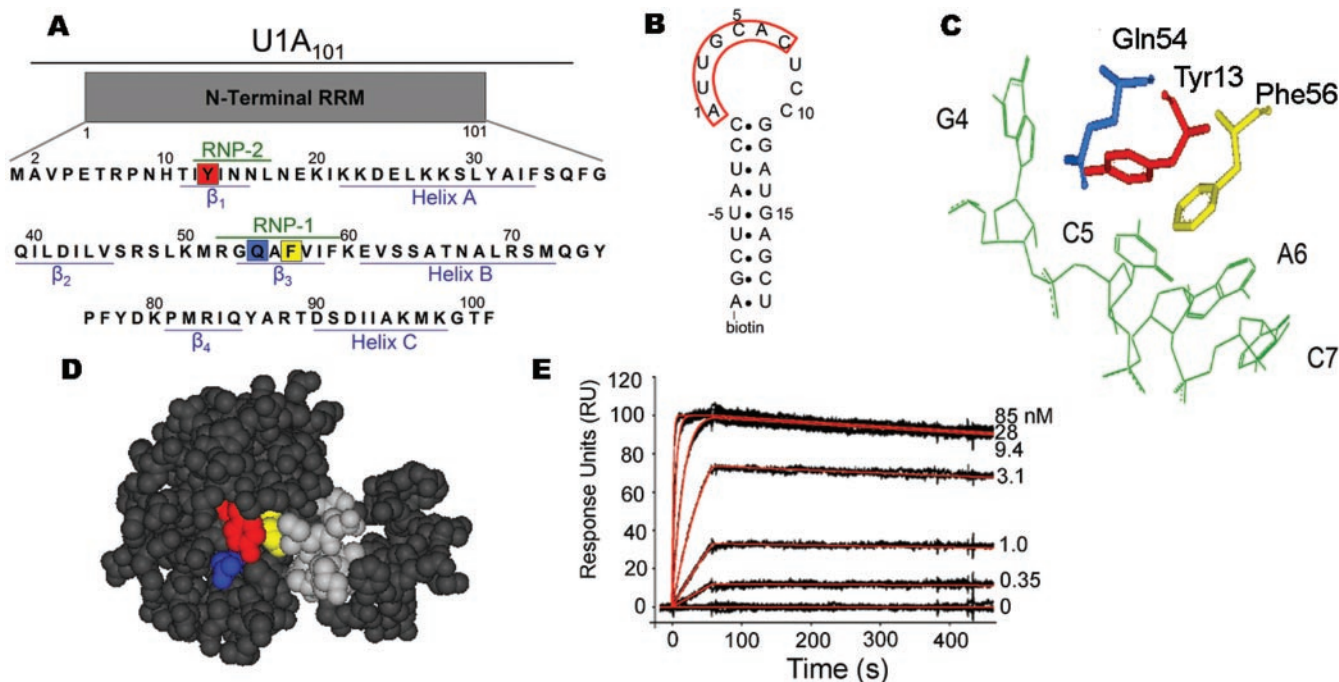
\*To whom correspondence should be addressed. Tel: +1 323 865 0655; Fax: +1 323 865 0158; Email: ilaird@usc.edu

Present address:

Phinikoula S. Katsamba, Center for Biomolecular Interaction Analysis, School of Medicine, University of Utah, Salt Lake City, UT 84132, USA

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oupjournals.org



**Figure 1.** Representation of protein, RNA target and the RNA–protein interaction. (A) The amino acid sequence of RRM1 of U1A is indicated, with Tyr13, Gln54 and Phe56 highlighted, and secondary structure features indicated. (B) U1hpII RNA used for our kinetic studies. Nucleotides U–5 to G15 are identical to the natural sequence. The numbering scheme is based on numbering of the loop residues 1–10, with backward and forward numbering of 5' and 3' stem residues, respectively. Key loop residues 1–7 have been highlighted. The 5' A carries a biotin. (C) Structural representation of stacking interactions between Gln54 (blue), Tyr13 (red) and Phe56 (yellow) and bases G4, C5 and A6, respectively, based on the U1A/U1hpII co-crystal structure (9). The RNA is shown in green. (D) Space-filling representation of the free U1A protein based on the solution structure of an amino acid 2–117 fragment (30). Tyr13 (exposed, red), Phe56 (partially hidden, yellow) and Gln54 (blue) are indicated. The C-terminal helix region is highlighted in light gray. (E) Sensorgram showing the interaction of wild-type U1A RRM1 (amino acid 1–101) with the U1hpII sequence shown in (B). A 1 min association phase was followed by a 7 min dissociation. Binding data are indicated in black; the interaction model, based on kinetic analysis, is marked in red. Increasing protein concentrations were run in triplicate and in random order over the RNA surface at the concentrations indicated. Kinetic parameters for the experiments are shown in Table 1.

protein–RNA complexes (9–16). The two key aromatic residues are the one in RNP-2 and the central aromatic residue in RNP-1; they are both involved in aromatic stacking interactions with RNA bases in all co-complexes examined to date (9–16). Of the remaining two aromatic residues, one resides near the RNA-binding surface (in the case of U1A, this is, instead, a glutamine) and probably functions as a hydrophobic spacer between nearby residues, and the other lies at the back of the RRM, distal to the RNA. The stacking interactions of the two RNA-binding aromatic residues do not appear to be sequence-specific, since all four bases have been observed to stack onto the conserved phenylalanine or tyrosine residues (9–16). Although they appear to provide little specificity, these interactions contribute importantly to the strength of RNA binding, as evidenced by the substantial loss of affinity observed when these residues are individually mutated (17–22). The importance of the two conserved aromatic residues for RNA binding has been well established by studies of a variety of proteins, using biochemical and biophysical approaches (9–16). However, the kinetic roles of these residues have not been well studied, and their relative contributions to mediating association with the RNA on the one hand, and maintaining complex stability on the other, remain unclear.

Here, we investigate the kinetic role of the aromatic residues in detail, using U1A, the A protein of the U1 snRNP particle, as a model system. U1A contains two RRM domains, but all current evidence suggests that the C-terminal domain is not

required for RNA binding (23–25). The N-terminal RRM of U1A (RRM1, herein also referred to as ‘U1A’) binds with picomolar affinity to U1 hairpin II (U1hpII), a stem–loop in the U1 snRNA, thereby recruiting U1A into the U1snRNP complex (24). Of the 10 loop nucleotides, 7 are highly conserved and are critical for tight binding of U1A, as demonstrated by *in vitro* selection and biochemical experiments (Figure 1B) (26–28). U1A also exhibits an autoregulatory activity: it prevents polyadenylation of its own message by binding to a structure in the 3'-untranslated region of its mRNA (29) (the polyadenylation inhibition element, which resembles a fused duplicate U1hpII structure and binds two copies of U1A). U1A RRM1 is the most widely studied RRM domain and has been used as the paradigm for high-affinity RNA binding by a single RRM.

Crystallographic and NMR studies have identified Tyr13 and Phe56 as the two key aromatic residues that stack onto bases of the RNA targets of U1A (in the case of U1hpII, onto loop nucleotides C5 and A6, respectively) (Figure 1C) (9,13,30). In the free protein, both Tyr13 and Phe56 lie on the RRM surface (Figure 1D). Depending on the fragment of U1A analyzed and the analysis conditions, they are either both solvent accessible (9,31) or Phe56 may be hidden by a C-terminal  $\alpha$ -helix (helix C, consisting of residues 91–98) (30). The exact positions of the two aromatic residues in the free protein do not appear to be completely static (9,30,31). However, in the RNA-bound form, Tyr13 and Phe56 appear to

be locked into place and form the nexus of a network of interactions that hold the RNA on the RRM (Figure 1C) (9,13). Phe56 is stacked onto loop base A6, which in turn is stacked onto C7, which in turn is stacked onto Asp92, a residue located in the C-terminal  $\alpha$ -helix (9). Thus, through the RNA, Phe56 connects to helix C, which is thought to clamp down on the bound RNA (13). Tyr13 links to another part of the protein that adjusts during RNA binding: the loop between the two central  $\beta$ -strands ( $\beta$ 2– $\beta$ 3 loop, residues 46–52, Figure 1A). In the bound complex, Tyr13 stacks onto C5, and its position is stabilized by a strong hydrogen bond from its hydroxyl group to the side chain carbonyl of Gln54. Gln54 in turn contacts several other residues. Its side chain amine makes hydrogen bonds to the main chain carbonyls of Lys50 and Arg52. Lys50 and Arg52 lie in the flexible  $\beta$ 2– $\beta$ 3 loop region, which becomes ordered upon RNA binding and partially protrudes through the RNA loop, playing a critical role in the stability of the U1A/U1hpII complex. Arg52 interacts with A1 in the RNA loop, as well as the closing C:G base pair at the top of the stem. The Gln54 side chain also stacks on base G4. Thus, Phe56 and Tyr13, and the interacting Gln54, are at the center of direct and indirect interactions with five of the seven highly conserved loop nucleotides (Figure 1B) as well as with the closing base pair at the top of the stem. Their key role is exemplified by the >1000-fold loss in affinity resulting from mutation of Tyr13 or Phe56 to non-aromatic residues, or Gln54 to Glu or Asn (7,17,18,20,22,32). The effects of these mutations on the dynamics of RNA binding have not been examined systematically.

In investigating the kinetic effects of mutation of Tyr13, Phe56 and Gln54, we were particularly interested in dissecting the role of these residues in complex association and/or dissociation. Using mutational analysis, kinetic studies and salt dependence experiments, we have previously shown that the positively charged residues Lys20, Lys22 and Lys50 help recruit the RNA through electrostatic interactions (7). This 'lure' step is thought to be followed by the formation of close-range interactions, as the conformations of the RNA and protein adapt in an induced fit model. We wanted to investigate whether it might be possible to dissect which of these close-range interactions occur very early in complexation, and thus might play a role in association. We utilized a surface plasmon resonance-based biosensor (Biacore) for our studies, as it provides high-quality data describing the kinetics of RNA–protein interactions (Figure 1E) (33). Our results suggest that Tyr13, Phe56 and Gln54 each play a role in both association and complex stability, but that the nature of substitutions made at these positions determines the extent of the deleterious effects. By combining our kinetic data with molecular dynamics simulations of wild-type and mutated proteins in their free and RNA-bound forms, we are able to provide structural correlates for the binding data. Based on our results, we propose a tentative model for the establishment of close-range interactions during U1A/U1hpII complex formation.

## MATERIALS AND METHODS

### Construction of U1A mutants and protein purification

Throughout these studies, an N-terminal fragment of the human U1A protein (amino acids 1–101, herein referred to

as U1A) containing the first RRM was used (7). This fragment has been demonstrated to be necessary and sufficient for specific and high-affinity binding to U1hpII (24,25). The U1A fragment was inserted into a modified pET3d vector such that a Myc and a (His)<sub>6</sub> tag were appended to the C-terminus of the RRM, as described previously (7). For cloning purposes, engineered restriction sites were introduced within the U1A coding region. All clones were generated by digestion of restriction sites that flank the area to be mutated and replacement with complementary oligonucleotides encoding the desired substitutions or deletions. The sequence identity of each clone was confirmed using both restriction digests and sequencing. Proteins were expressed in *Escherichia coli* BL21(DE3) (Novagen, Madison, WI, USA) and purification was carried out using the hexahistidine tag at the C-terminus of the protein (7,34). After binding to Ni<sup>2+</sup> beads (Qiagen, Valencia, CA, USA) samples were eluted using increasing concentrations of imidazole (50–250 mM). The concentration of each protein was estimated using the Bradford assay (BioRad, Hercules, CA, USA) and confirmed by Coomassie blue staining of an extensive protein dilution series next to a standard on SDS–PAGE gels.

### Biosensor analysis

Binding experiments were performed on a BIACORE 2000 instrument (Biacore Inc., Piscataway, NJ, USA). U1hpII RNA was chemically synthesized carrying a 5'-biotin tag (Dharmacon Research, Boulder, CO, USA) to allow immobilization of the RNA onto streptavidin-coated sensor chips (SA chips, Biacore Inc.). RNA was diluted to a final concentration of 1  $\mu$ M in HBS buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% surfactant P20) followed by heating at 80°C for 10 min and cooling to room temperature to allow annealing of the stem. The sample was then diluted 500-fold in running buffer (10 mM Tris–HCl, pH 8.0, 150 mM NaCl, 5% glycerol, 125  $\mu$ g ml<sup>-1</sup> tRNA, 1 mM DTT, 0.05% surfactant P20) and injected over the sensor chip surface at 10  $\mu$ l min<sup>-1</sup> at 20°C. (We have recently removed BSA from this buffer because it was causing surface degradation problems. Its removal did not appear to cause increased background signal problems.) To provide an optimal comparison of the results obtained from all different U1A mutants, we prepared an intermediate density RNA surface (100–125 resonance units) that would yield sufficient signal even when proteins with lower affinities were used. In several cases, the proteins were also analyzed using a higher density surface (~300 resonance units), which yielded comparable results. To test for the specificity of the RNA-binding interaction, binding of all proteins to a control surface consisting of a U1hpII RNA in which the order of the loop nucleotides had been reversed from 5'-AUUGCACUCC-3' to 5'-CCUCACGUUA-3' ('reverseU1hpII') was also assessed. Reversion of the loop sequence changes 8 of the 10 loop nucleotides, including 6 of the 7 highly conserved loop residues (17,26,28,35) but leaves the loop structure intact. Proteins were serially diluted in running buffer to the concentrations indicated in Figures 1 and 2 and injected at 20°C at a flow rate of 50  $\mu$ l min<sup>-1</sup> for 1 min. Disruption of any complex that remained bound after a 5 min dissociation was achieved using a 1 min injection of 2 M NaCl at 20  $\mu$ l min<sup>-1</sup>. Samples with different concentrations of

protein were injected in random order, and every injection was performed in triplicate within each experiment. All experiments were done three to six times. In order to subtract any background noise from each dataset, all samples were also run over an unmodified sensor chip surface and random injections of running buffer were performed throughout every experiment ('double referencing'). Data were processed using Scrubber (developed by the Biomolecular Interaction Facility at the University of Utah, [www.cores.utah.edu/interaction](http://www.cores.utah.edu/interaction)) and analyzed using CLAMP (36) and a simple 1:1 Langmuir interaction model with a correction for mass transport (37). The results for all mutants were compared (to the wild-type protein and to each other) using the Student's *t*-test to determine whether or not they were statistically significant.

### Molecular dynamics simulations

Isobaric molecular dynamics simulations of hydrated proteins and hydrated protein–RNA complexes were run on a Silicon Graphics Challenge computer using AMBER version 6.0 (38). The simulations were run in parallel mode; simulations of the free protein were run for 900 ps using four nodes and the protein–RNA simulations were run for 1 ns on six nodes. Free protein simulations were performed on the wild-type and three mutated proteins, Gln54Ala, Gln54Glu and Gln54Asn, beginning from the first model/ensemble member of the NMR solution structure of the N-terminal RRM of the U1A protein (30) obtained from the Protein Data Bank (PDB ID: 1FHT). The 116-residue NMR structure of the protein was chosen as a starting point rather than the (shorter) crystal structure (PDB ID: 1IOA) because it contains the full C-terminal helix (residues 91–98). Protons were added to the structure in the LEaP module of AMBER 6.0 (39), and all atoms of the protein were then minimized *in vacuo* for 1500 steps (500 steps of steepest decent, followed by 1000 steps of conjugate gradient minimization). The minimized structure was then inserted into the center of a periodic box containing TIP3P water molecules. Solute atoms were at least 9 Å from the boundary; this resulted in a box with dimensions 58 × 75 × 59 Å. Water molecules closer than 1 Å to any solute atom were deleted. The wild-type simulation contained 6234 water molecules and a total of 20 649 atoms, including 1947 protein atoms. All simulations were run using the SANDER module of AMBER 6.0 and SHAKE (40) was applied to all hydrogen atoms. Equilibration of the solvent molecules was achieved by first raising the temperature of the system to 298K during the first 10 000 steps (20 ps) with position-restraint of all protein atoms with a force constant of 20 kcal/(mol Å). The solute atoms remained so constrained for another 40 000 steps, allowing the water to relax around the solute at 298K. After this equilibration period, all subsequent simulations were run using the interpolated particle mesh Ewald method to determine Lennard–Jones and electrostatic interactions (41). Following the 100 ps solute-restrained period, the restraint on the solute atoms was removed and a 900 ps simulation was performed, the first 100 ps of which was considered to be part of the equilibration of the system. The target pressure was 1 atm, the time constant was 0.002 ps and the Lennard–Jones cutoff was 8 Å.

To study the U1A/U1hpII complex, 1 ns simulations were carried out on the complex of the wild-type protein and on

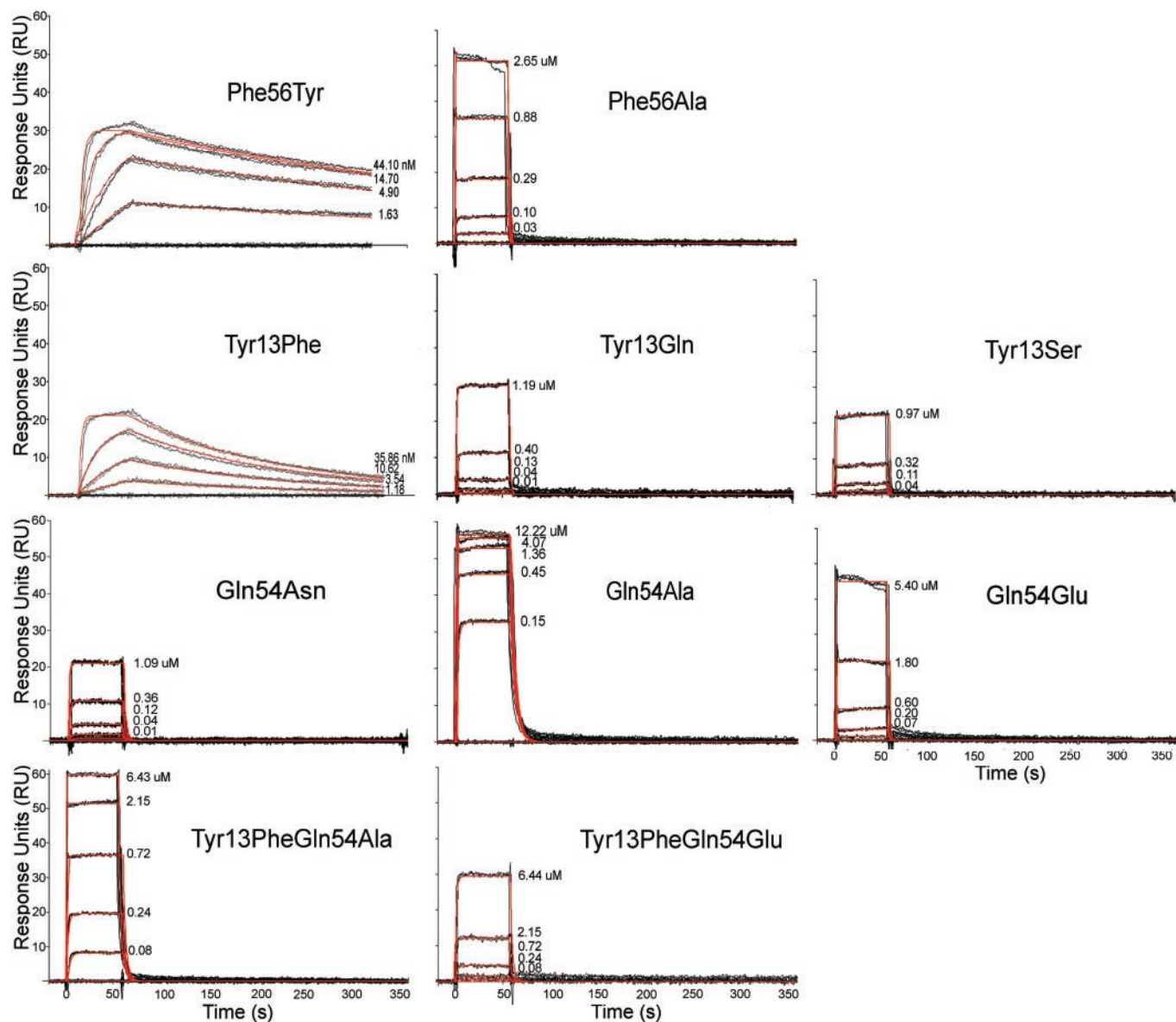
those of three mutated proteins, Gln54Ala, Gln54Glu and Gln54Asn. Calculations were based on the B and Q chains of the X-ray coordinates of human U1A (amino acids 2–97) complexed with the RNA hairpin 5'-AAUCCAUGCACUCCGGAUUU-3' (9) (PDB ID: 1URN). We removed the 5' adenine and two uracil bases and extended the stem with a 5 bp RNA helix, in order to better match the RNA used in the Biacore experiments: 5'-AGCUUAUCCAUGCACUCCGGAUUAAGCU-3'. The RNA stem extension was added by superimposing a 9 bp duplex RNA (built with the NUCGEN module in AMBER6) onto the experimentally determined stem, such that 4 bp of the 9 bp duplex were fitted to 4 bp of the experimental structure. The four fitted base pairs of the ideal duplex were then deleted, leaving the 5 bp extension. Since the protein of the X-ray structure was incomplete, it was necessary to build side chains for Lys20 and Lys96. In addition, the X-ray structure contained two mutated residues (His31 and Arg36), which were mutated back to the wild-type residues, Tyr31 and Gln36, respectively. The mutated protein residues and the RNA backbone atoms connecting the X-ray structure to the NUCGEN-built RNA stem extension were relaxed using a 3000-step minimization *in vacuo*, in which all other atoms were restrained. Water molecules present in the X-ray structure were retained for the simulation, except that the removal of 8 of these 157 water molecules was necessary to allow the positioning of the extended RNA stem and the sodium counterions (crystal water molecules closer than 1 Å to the atoms of the extended RNA stem were removed). Using LeaP, sodium ions were added to make the complexes electroneutral (22 ions for the Gln54Glu mutant and 21 for the wild-type, Gln54Asn and Gln54Ala complexes). The solvent equilibration and data accumulation simulations for the protein–RNA simulations followed the procedure outlined above for the free protein simulations. For the wild-type complex, the simulation system included 27 496 atoms and contained 8333 water molecules in a box of dimensions 72.5 × 59.1 × 81.9 Å.

The simulation data were analyzed as follows. Individual atom interactions between protein and RNA atoms were identified using the analysis algorithm PRORNA (E.J. Chambers, M.J. Law, K.A. Patel, M.Z. Bayramyan and I.S. Haworth, manuscript in preparation). Other analyses were performed using PTRAJ in AMBER 6 and MOLTOOL (I.S. Haworth, unpublished data). Simulation dynamics were visualized using VMD (42) and individual structures were also visualized with WebLab Viewer Pro 3.7 (Molecular Simulations Inc., Copyright 2000. San Diego, CA, USA).

## RESULTS AND DISCUSSION

### Aromatic stacking at position 56 is critical for complex stability

We have previously examined the RNA-binding kinetics of a Phe56Ala RRM1 mutant of U1A (using a fragment encompassing amino acids 1–101), and observed a dramatic increase in the dissociation rate, whereas the association rate appeared relatively unaffected (7). When we reanalyzed this mutant in the process of comparing its behavior with that of other aromatic substitutions, we confirmed that the large loss in affinity (25 000-fold) was due predominantly to a severe loss in



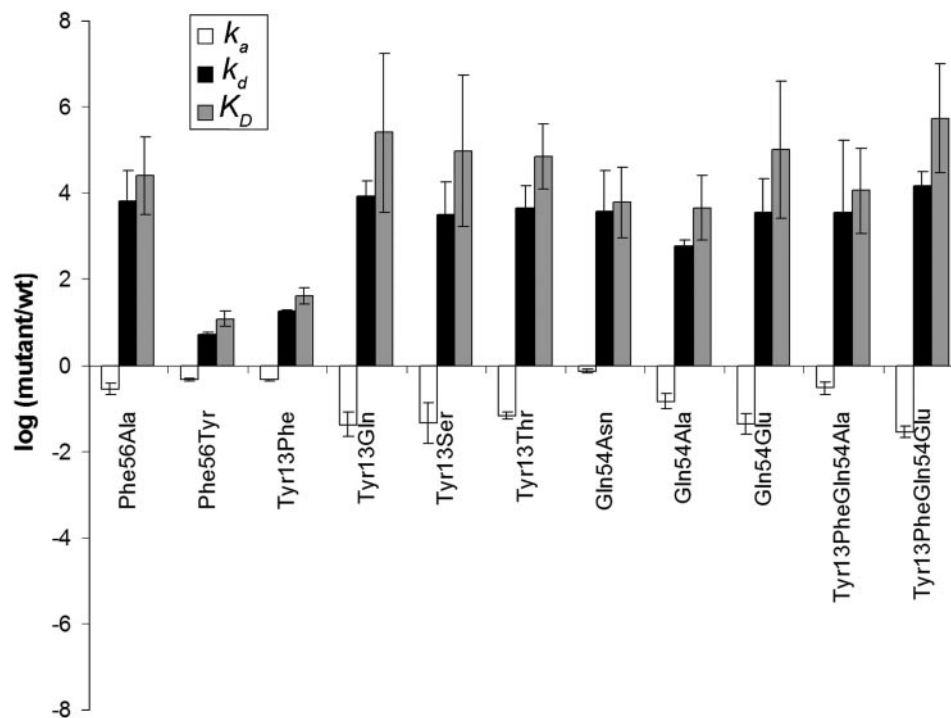
**Figure 2.** Sensorgrams showing kinetic analyses of U1A mutants interacting with U1hpII. The different mutations made at each position in the protein are shown from left to right in each row. For visual uniformity, the *x*- and *y*-axes are identical in each sensorgram. As the Tyr13Ser and Tyr13Thr plots were very similar, in the interest of space only Tyr13Ser is shown. Increasing protein concentrations were run in triplicate and in random order over the RNA surface, at the concentrations indicated. Association was monitored for 1 min, followed by a 5 min dissociation. The black lines represent triplicate protein injections, and the red lines represent the global fit of the datasets using CLAMP (36). Kinetic parameters for the experiments are shown in Table 1.

complex stability (over three orders of magnitude;  $P < 0.01$ ) (Figures 2 and 3, Table 1), (7,22), with a 4-fold loss in the association rate ( $P < 0.001$ ), similar to that previously observed. The effect on dissociation was 4-fold more than we had seen previously, perhaps owing to minor differences in experimental conditions. For comparison, we examined a mutant in which Phe56 had been replaced by Tyr, thus maintaining the aromatic nature of the position. This mutant showed a very modest ( $\sim 12$ -fold) loss in affinity (Figures 2 and 3, Table 1), consistent with previous reports (20,22,32). The weakened binding of Phe56Tyr was due to a small but significant ( $\sim 2$ -fold;  $P < 0.001$ ) effect on association and a slightly larger increase in dissociation rates ( $\sim 5$ -fold;  $P < 0.001$ ), suggesting that the presence of tyrosine in this

position is only mildly disruptive. This is probably caused by the introduction of the hydroxyl, which may affect the position of surrounding amino acids; replacement of Phe56 with Trp in two other studies showed no or a negligible effect on RNA binding (22,32). Taken together, our results support a critical role for an aromatic residue at position 56 in complex stability, and a very minor role in association.

#### **Tyrosine 13 is crucial for complex stability and has a possible role in association**

We next examined the effects of mutating Tyr13, which lies on the RNA-binding surface in the unbound RRM, but, in contrast to Phe56, appears always to be solvent exposed, even in the



**Figure 3.** Comparison of association and dissociation rates and affinity for wild-type and mutated U1A. To visualize relative differences accurately, we plotted the logarithm of the mutant over wild-type values. Error bars indicate the standard error.

**Table 1.** Kinetic parameters for U1hpII interaction with U1A and U1A mutants<sup>a</sup>

Protein	$k_a$ ( $M^{-1} s^{-1}$ )	$k_a$ WT/Mut	$k_d$ ( $s^{-1}$ )	$k_d$ Mut/WT	$K_D$ ( $=k_d/k_a$ ) (M)	$K_D$ Mut/WT
U1A	$9.5 \pm 0.1 \times 10^6$	1	$3.2 \pm 0.4 \times 10^{-4}$	1	$3.4 \pm 0.4 \times 10^{-11}$	1
Phe56Ala	$2.7 \pm 0.6 \times 10^6$	4	$2.1 \pm 0.4 \times 10^0$	6500	$8 \pm 1 \times 10^{-7}$	25 000
Phe56Tyr	$4.4 \pm 0.5 \times 10^6$	2	$1.7 \pm 0.1 \times 10^{-3}$	5	$4.0 \pm 0.8 \times 10^{-10}$	12
Tyr13Phe	$4.3 \pm 0.2 \times 10^6$	2	$5.8 \pm 0.1 \times 10^{-3}$	18	$1.3 \pm 0.1 \times 10^{-9}$	40
Tyr13Gln	$4.0 \pm 0.8 \times 10^5$	24	$2.8 \pm 0.2 \times 10^0$	8700	$8 \pm 3 \times 10^{-6}$	250 000
Tyr13Ser	$4.3 \pm 1.5 \times 10^5$	22	$1.0 \pm 0.2 \times 10^0$	3200	$3 \pm 1 \times 10^{-6}$	100 000
Tyr13Thr	$6.4 \pm 0.6 \times 10^5$	14	$1.4 \pm 0.1 \times 10^0$	4400	$2 \pm 1 \times 10^{-6}$	60 000
Gln54Ala	$1.4 \pm 0.3 \times 10^6$	7	$1.9 \pm 0.1 \times 10^{-1}$	600	$1.5 \pm 0.3 \times 10^{-7}$	4500
Gln54Glu	$4.2 \pm 0.8 \times 10^5$	23	$1.2 \pm 0.3 \times 10^0$	3600	$3 \pm 1 \times 10^{-6}$	100 000
Gln54Asn	$6.9 \pm 1.7 \times 10^6$	1	$1.2 \pm 0.3 \times 10^0$	3600	$2.1 \pm 0.4 \times 10^{-7}$	6000
Tyr13PheGln54Ala	$2.8 \pm 0.8 \times 10^6$	3	$1.2 \pm 0.5 \times 10^0$	3600	$3.8 \pm 0.8 \times 10^{-7}$	11 000
Tyr13PheGln54Glu	$2.7 \pm 0.2 \times 10^5$	35	$4.8 \pm 0.4 \times 10^0$	15 000	$1.9 \pm 0.3 \times 10^{-5}$	500 000
Tyr13SerGln54Ala	No binding	—	No binding	—	No binding	—
Tyr13SerGln54Glu	No binding	—	No binding	—	No binding	—
Tyr13GlnGln54Ala	No binding	—	No binding	—	No binding	—
Tyr13GlnGln54Glu	No binding	—	No binding	—	No binding	—

<sup>a</sup>All differences in  $k_a$ ,  $k_d$  and  $K_D$  between wild-type U1A and the mutants were statistically significant ( $P < 0.05$ , Student's *t*-test), except for the  $k_a$  of Gln54Asn, which was not significantly different from that of the wild-type protein ( $P > 0.2$ ). The SEM is indicated in the  $k_a$ ,  $k_d$  and  $K_D$  columns. Statistical significance of differences between different mutants is indicated in the text. The fold difference between wild-type and mutant values for  $k_a$ ,  $k_d$  and  $K_D$ , respectively, is given in columns to the right of the values.

presence of helix C (Figure 1D). Thus, Tyr13 is a good candidate site for the formation of close-range interactions early in complex assembly. Previous work by other laboratories has shown that a mutation of Tyr13 to Phe (in the context of a 1–101 RRM1 fragment) causes a 100-fold loss in affinity (17,18). This mutation removes the hydroxyl that bridges to Gln54 and might thereby affect the interaction of other parts of the protein with the RNA. In an attempt to discriminate between the effects of the aromatic stacking contribution and those of the Tyr13 hydrogen bond, we replaced Tyr13

with Phe, Gln and Ser. Replacement with Phe led to a ~40-fold reduction in affinity, owing to a 2-fold loss in association rate ( $P < 0.001$ ) and a ~20-fold destabilization of the complex ( $P < 0.001$ ) (Figures 2 and 3, Table 1). When comparing the Phe56Tyr and Tyr13Phe substitutions, there was no significant difference in the  $k_a$  ( $P > 0.8$ ), indicating that the identity of the aromatic residue at these positions is not a key factor in association. However, the difference in  $k_d$  was highly significant ( $P < 0.001$ ), supporting the idea that the loss in stability of the Tyr13Phe complex is due to the absence of the hydroxyl bond

to Gln54, which would normally stabilize the stacking interaction at position 13, as well as help position the  $\beta 2$ – $\beta 3$  loop in the RNA loop (43).

Removal of the aromatic side chain of Tyr13 was very disruptive and led to a dramatic loss in affinity. Neither a Gln nor a Ser could compensate for loss of the Tyr side chain (Figures 2 and 3, Table 1), although the Tyr13Ser complex appeared to be slightly more stable than the Tyr13Gln complex. A similar inability of Thr to replace Tyr13 was observed by Hall and coworkers (17). To verify this result, we examined the kinetics of a Tyr13Thr mutant and observed binding behavior very similar to Tyr13Ser (no significant differences in  $k_a$ ,  $k_d$  or  $K_D$  compared with Tyr13Ser, Figure 3, Table 1). In combination with the data for Tyr13Phe, this suggests that the hydrogen bond from residue 13 to Gln54 contributes minimally to binding, in the absence of aromatic stacking at that position.

The >1000-fold increase in complex dissociation seen for Tyr13Gln and Tyr13Ser is similar to that seen for the Phe56Ala mutant (Figures 2 and 3, Table 1). Interestingly, both Tyr13Gln and Tyr13Ser showed a significantly larger loss in association rate than Phe56Ala ( $P < 0.03$ ), suggesting that the aromatic residue at position 13 is more important for association than the one at position 56. The slower association of the Tyr13 mutants compared with the Phe56 mutant could be due to slight perturbations of the RRM structure that might affect how its positive charges are presented to the RNA. However, it could also suggest a role for the aromatic side chain of Tyr13 in splaying out of the seven highly conserved loop bases on the RRM surface. Such a role might be related to the fact that Tyr13 appears accessible in the free protein, whereas Phe56 might be buried (9,30,31). Although Phe56 appears solvent accessible in a crystal structure of the 2–95 fragment of the unbound protein (44), it is hidden in a solution structure of a 2–117 U1A fragment (Figure 1D) (30). A recent crystal structure of the longer fragment showed the C-terminal helix (residues 91–98) positioned away from the RRM surface (in the location it would normally occupy in the RNA-bound form of the protein), leaving Phe56 open to interactions (31). Although the location of the helix C in the latter structure could be due to the crystallization conditions (as suggested by the authors), it might simply be indicative of two alternative positions that this part of the protein is able to assume, as suggested by molecular dynamics simulations (45). Because the simulations did not show transitions between the two positions of helix C, it seems likely that other interactions precede the stacking of Phe56 on A6, and that these other interactions drive the rearrangement of helix C.

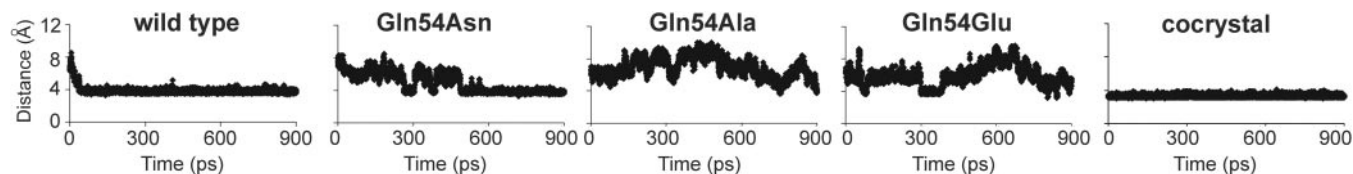
### Gln54 plays a key role in association and complex stability

Not only does Gln54 make a strong hydrogen bond to Tyr13 in the RNA–protein complex; it also stacks on G4 and helps position Lys50 and Arg52 as the  $\beta 2$ – $\beta 3$  loop rearranges to protrude through the RNA loop (9,30). Mutation of Gln54 to Ala or Glu severely inhibited RNA binding, weakening the affinity by three to five orders of magnitude (Figures 2 and 3, Table 1). The Gln54Glu mutation exhibited a  $\sim 4000$ -fold loss in complex stability ( $P < 0.006$ ), as well as a  $\sim 20$ -fold loss in association rate ( $P \ll 0.001$ ) (Figures 2 and 3, Table 1).

This loss in association might be explained by a role of Gln54 early in complex formation through any of its interactions with its amino acid neighbors and/or through its role in positioning G4. However, the association defect of the Gln54Ala mutant was less pronounced [7-fold loss in association ( $P < 0.008$ ), Figures 2 and 3, Table 1], indicating that Gln54Glu has an added association defect compared with Gln54Ala (the  $\sim 3$ -fold slower association rate of Gln54Glu versus Gln54Ala is significant;  $P < 0.05$ ). Gln54Glu also dissociated faster from the RNA than Gln54Ala, indicating that Glu at position 54 also interferes with complex stability ( $P < 0.02$ ) (Table 1). A conservative substitution, Gln54Asn, had been reported by Hall and coworkers to bind weakly (32). When we examined the binding kinetics of this conservative mutation, it became clear that the loss of affinity of this mutant derives entirely from a destabilization of the RNA–protein complex (Figures 2 and 3, Table 1). Strikingly, there was no difference in association rate between this mutant and wild-type protein ( $P > 0.05$ ). Thus, our results suggest that (i) mutation of the Gln side chain at position 54 is very deleterious to complex stability, particularly when replaced by Glu or Asn, and (ii) the more pronounced association loss observed with Gln54Glu compared with Gln54Ala is probably the result of negative effects of the Glu substitution on the presentation of the protein to the RNA, in addition to the loss of Gln interactions. Insight into the role of Gln at position 54 in both association and complex stability might be obtained by examining the structural consequences of mutations at this position, in the context of both the free protein and the complex. Thus, we undertook molecular dynamics simulations of the free protein and the protein–RNA complex.

### The Gln54Glu and Gln54Ala substitutions affect presentation of charge on the protein surface

We first carried out simulations of the wild-type U1A protein and proteins carrying Gln54Ala, Gln54Glu and Gln54Asn mutations in the absence of RNA, to examine whether structural rearrangements could explain differences in their association rates with U1hpII. In particular, we sought to understand the role of the glutamate mutation at position 54, which strongly disrupts the association with the RNA. Molecular dynamics simulations of the four proteins were based on a structure comprising amino acids 2–117 (PDB ID: 1FHT) (30) and were carried out for 900 ps. Broadly, the four simulations were structurally similar, with no major internal rearrangements of the free protein, suggesting that the association defects seen in the Ala and Glu mutants derive in large part from the inability of the Ala and Glu side chains to initiate the interactions normally made by Gln54. The conservative Asn substitution, which showed no association defect, would still be able to support these interactions sufficiently to mediate association successfully, although these interactions cannot be completed, resulting in loss of complex stability (see below). An important difference in the simulations is associated with the formation of a key interaction between Arg47 and Arg52: in both the wild-type protein and Gln54Asn, Arg52(N1[H]) establishes a stable hydrogen bond to the Arg47 backbone carbonyl oxygen (Figure 4, left two panels). However, this behavior is not seen in the simulations of the Gln54Ala and Gln54Glu mutants (Figure 4, third and fourth



**Figure 4.** Interaction between Arg52 and Arg47 in the wild-type U1A and proteins carrying Gln54Asn, Gln54Ala or Gln54Gln mutations. Molecular dynamics simulations of the proteins were based on the solution structure of an amino acid 2–117 fragment (30). The distances of Arg52(C $\gamma$ ) to the Arg47 backbone carbonyl were plotted for all four free proteins and the U1hpII/wild-type U1A complex (right panel). Note that a stable hydrogen bond forms in the wild-type protein and Gln54Asn, and is present in the RNA–protein complex. However, it is absent in the Gln54Ala and Gln54Glu mutants.

panel). Importantly, this interaction is also observed in the co-crystal structure of the U1A/U1hpII complex (9), and remains stable throughout a 900 ps simulation of the complex (Figure 4, right panel). Thus, it appears that the ‘head-to-tail’ Arg52–Arg47 arrangement, which is formed in the wild-type and Gln54Asn proteins but not in the Gln54Ala and Gln54Glu mutants, essentially locks the Arg52 side chain into a position in which it can receive the incoming RNA. Examination of the 3D structure of the four proteins shows Arg52 positioned similarly in the wild-type and Gln54Asn proteins, approximately equidistant from Lys20 and Lys22, which we have previously shown are important for the electrostatic approach of RNA and protein (7). In contrast, in the Ala and Glu mutants, Arg52 projects at a very different angle with respect to the RRM domain, and the distance to Lys20 and Lys22 is increased. The amino acid at position 54 does not appear to exert a direct effect on the behavior of the Arg52–Arg47 interaction through contact with either arginine. However, the structural rearrangement required for repositioning of Arg47 relative to Arg52 may be influenced by the flexibility of the  $\beta$ 2– $\beta$ 3 and  $\beta$ 1– $\alpha$ 1 loops, and this in turn is dependent on interactions between the  $\beta$ 2 and  $\beta$ 1 strands. Position 54 is located at the terminus of the  $\beta$ 2 strand, and an interaction develops between this position and Asn15 only in the wild-type and Gln54Asn simulations. Hence, side chain modification at position 54 may influence the association of  $\beta$ 2 and  $\beta$ 1, thereby altering the propensity for the Arg52–Arg47 interaction and (in the cases of Gln54Ala and Gln54Glu) incorrectly positioning the Arg52 side chain.

An additional factor resulting from the Arg52–Arg47 arrangement is an increase in the net positive charge in this crucial area of the protein, perhaps facilitating RNA recruitment. A negative charge in this area would probably be deleterious, potentially repelling the RNA, and this may well be the cause for the additional 3-fold loss in association seen with the Gln54Glu mutant compared with Gln54Ala. Examination of the 3D structure of the Gln54Glu mutant shows Glu54 positioned between Arg52 and Asn15, thus inserting negative charge between two amino acids that must establish key interactions with RNA bases. An interaction unique to the Gln54Glu mutant was also observed: a persistent hydrogen bond between Tyr13 and Glu54, in which Tyr13 is the H-bond donor. While it is possible that this might constrain Tyr13, which must be allowed to stack on loop base C5, this interaction is unlikely to be the cause of the Gln54Glu association defect; removal of the Tyr hydroxyl group (Tyr13Phe) does not alleviate the association defect of the Gln54Glu mutant (see discussion of double mutants, below). In addition, a hydrogen bond from Gln54 to Tyr13, though absent in the

free wild-type protein, is present in the RNA–protein complex (9). Thus, the association defect in the Gln54Ala and Gln54Glu mutants are probably caused by mispositioning of Arg52, and the added defect of Gln54Glu probably stems from the added negative charge. Based on previous simulations, Tang and Nilsson have proposed that Arg52 initiates close-range interactions following the electrostatic approach of RNA and protein (46), and the Arg52 side chain position and electrostatic environment are therefore likely to be highly relevant for proper association.

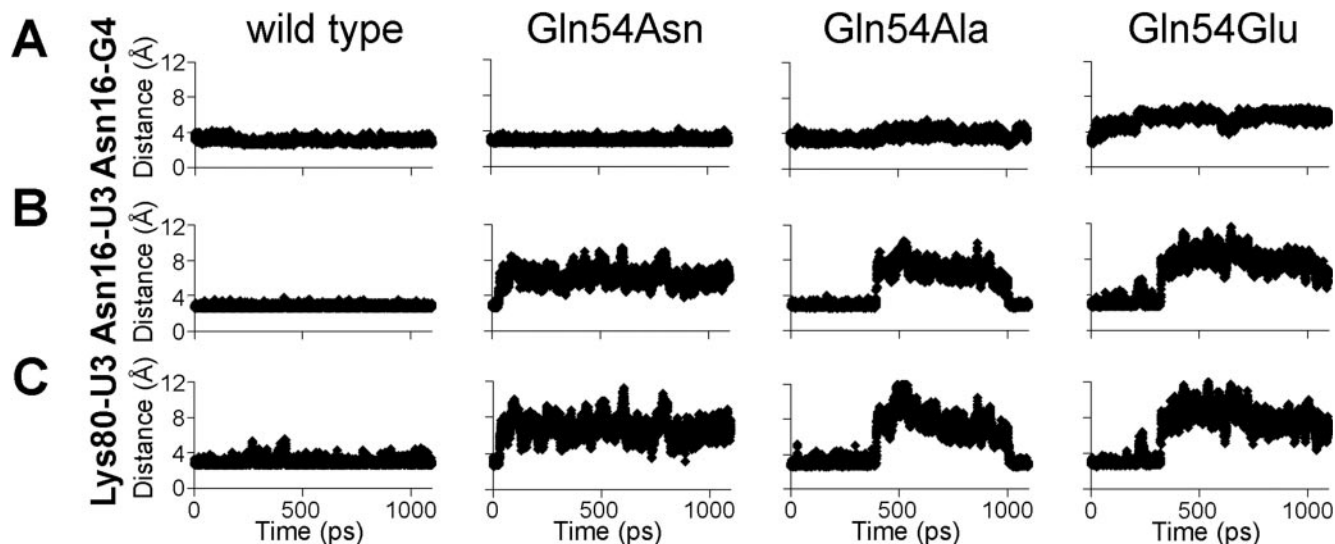
#### Mutation of Gln54 disrupts interactions of the protein with U3 and G4

To explore the basis for the instability of protein–RNA complexes containing Gln54 mutations, we also carried out molecular dynamics simulations of these complexes. The co-crystal structure of RRM1 with U1hpII (9) was modified by extending the RNA stem by 5 bp to more closely mimic the RNA target used for the kinetic studies (see Materials and Methods). In the simulation of the wild-type complex, Gln54 remained stacked with G4, whereas this stacking interaction was lost in simulations of the complexes with the Gln54Ala, Gln54Glu and Gln54Asn mutants. The loss of stacking with Ala, Glu or Asn at position 54 affected the positioning of G4, disrupting the hydrogen bond between G4(O6) and the backbone N[H] of Asn16 (Figure 5A). In turn, this appeared to affect the interaction between Asn16 and U3, as seen by the disruption of the hydrogen bond between Asn16(O $\delta$ ) and U3(N3[H]) (Figure 5B). This coincided with loss of the hydrogen bond between U3(O4) and Lys80(N $\zeta$ [H]) (Figure 5C). The disruption of hydrogen bonds to U3 and G4 was present for longer periods during the simulation of Gln54Glu and Gln54Asn than that of Gln54Ala, providing an explanation for the greater instability of the Gln54Glu/RNA and Gln54Asn/RNA complexes.

#### Aromatic side chain loss combined with Gln54 mutations abolishes binding

To better understand the kinetic effects of the interaction between Tyr13 and Glu54, we generated six mutants in which Tyr13Phe, Tyr13Gln and Tyr13Ser were combined with Gln54Ala or Gln54Glu. Of these, only the combinations carrying Tyr13Phe bound to RNA; apparently, a Gln54 mutation in addition to loss of the aromatic side chain at position 13 is devastating to RNA binding (Figures 2 and 3, Table 1). This effect has been observed previously and has been attributed to the local cooperative nature of the interactions between Tyr13, Phe56 and Gln54 (32). The Tyr13Phe replacement did not





**Figure 5.** Molecular dynamics simulations of wild-type, Gln54Asn, Gln54Ala and Gln54Glu U1A proteins complexed with U1hpII RNA. Molecular dynamics simulations of the complexes were based on the U1A/U1hpII co-crystal structure (9), with an extended RNA stem included to better match the Biacore experiments. In the wild-type complex, an important stacking interaction between Gln54 and G4 allows G4(O6) to hydrogen bond stably to the backbone N[H] of Asn16 (A), and in turn hydrogen bond interactions between Asn16-U3 (B) and U3-Lys80 (C) are stabilized. In the mutants, a concerted disruption of the hydrogen bonds between Asn16 and G4 (A), Asn16(side chain O) and U3(N<sub>3</sub>[H]) (B), and the Lys80(N $\zeta$ [H]) and U3(O<sub>4</sub>) (C), is seen.

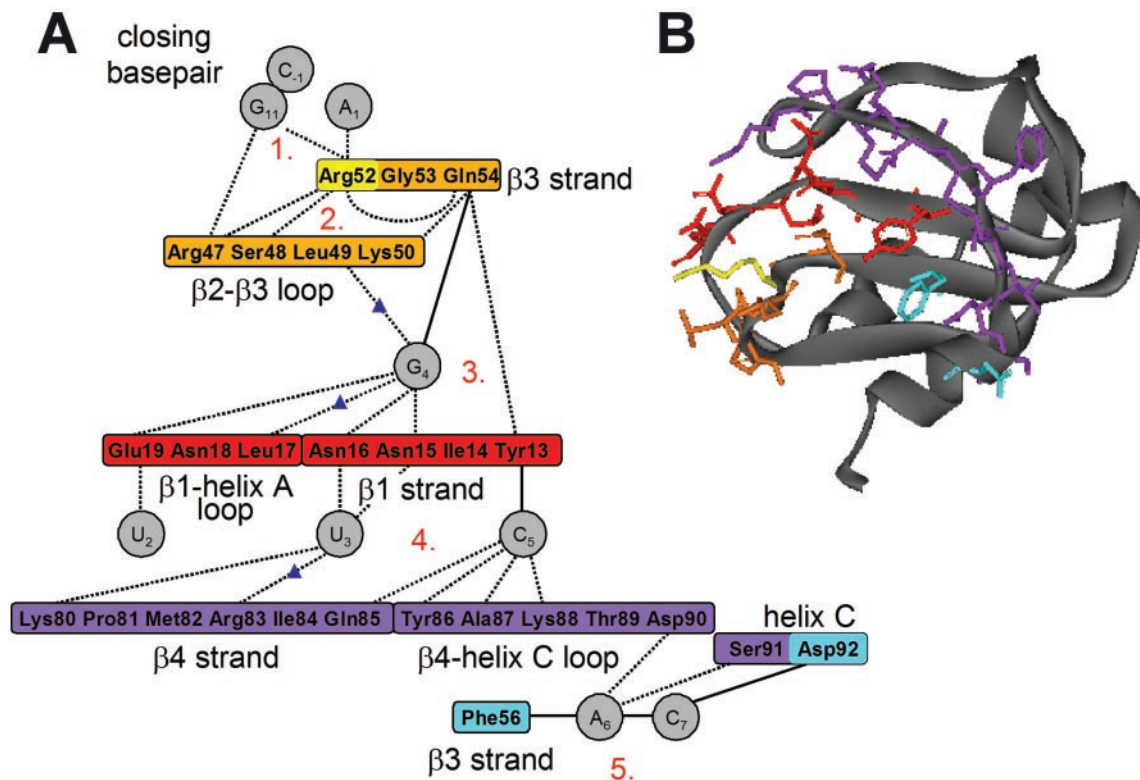
significantly alter the association rate of the two Gln54 mutants ( $P > 0.05$ ). Replacement of Tyr13 with Phe destabilized both the Gln54Ala and Gln54Glu bound complexes by 4- to 6-fold, hinting that loss of the hydroxyl group may result in a small additional destabilization, although for the Gln54Ala mutant the difference in  $k_d$  was not significant. The six mutants reinforce the idea of cooperativity of Tyr13 and Gln54 interactions with RNA (32). In addition, the inability of the Tyr13Phe mutation to rescue the Gln54Glu mutation indicates that the Tyr13-Glu54 hydrogen bond in the Gln54Glu mutant is not the cause of its poor binding ability.

### A multi-step model for complex formation

Binding of RNA to proteins has been suggested to occur according to an induced fit mechanism, in which both partners adapt their conformations during complex formation (47). That this occurs during the U1A/U1hpII interaction is evident from the differences in the structures of the free and bound molecules observed both experimentally (9,13,44) and theoretically (21,45,46,48,49). In the protein, the changes upon binding involve protrusion of the flexible loop between  $\beta$ -strands 2 and 3 (the  $\beta$ 2- $\beta$ 3 loop) through the RNA loop, and a moving away of the C-terminal helix, providing the RNA access to the full RRM surface, followed by a clamping down of the N-terminal end of helix C onto the RNA. In the RNA, the changes involve the splaying out of the loop bases to form sequence-specific interactions with the protein. Based on molecular dynamics simulations, Tang and Nilsson have suggested that binding occurs in three steps (46): first, electrostatic interactions bring the molecules together with the correct respective orientations; second, binding is initiated by early close-range interactions (proposed to occur via Arg52); and finally, simultaneous structural rearrangements of RNA and protein allow formation of the final complex. We have

previously shown that mutation of Lys20 + Lys22 or Lys50 predominantly slows association, whereas mutation of residues on the RNA or protein responsible for close-range interactions predominantly affects complex stability (7). These observations support the existence of steps 1 and 3 in the Tang and Nilsson model (46). To clarify the sequence of events following positioning of both molecules, we examined the kinetic effects of mutations of Tyr13, Phe56 and Gln54. We found that the magnitude of the association defect differs for different residues and for distinct substitutions at each position.

What do the kinetic analyses tell us about the steps in association? In Figure 6, we propose a tentative sequential model for the formation of close-range interactions in the complex, based on our observations, the model proposed by Tang and Nilsson (46) and structural studies (9). Once the RNA and protein are aligned based on electrostatic attraction (7,46), the first close-range interaction could well be charge-based. An initiating role for Arg52 appears reasonable; it projects into the solvent, and one of the many hydrogen bonds formed by Arg52 in the complex is to G11 in the C:G base pair at the top of the RNA stem, which provides a stable target (Figure 6A, Step 1). The second contact of Arg52 to the RNA is a hydrogen bond to nearby loop base A1. The ensuing steps might result from additional hydrogen bonds that Arg52 makes to the main chain carbonyls of residues in the  $\beta$ 2- $\beta$ 3 loop and to Gln54 (Figure 6A, Step 2). Arg52 would thus be a key residue in association, initiating both close-range contacts with the RNA and the positioning of the  $\beta$ 2- $\beta$ 3 loop. Indeed, mutation of Arg52 to Gln has been reported to abolish RNA binding by U1A (44). The interaction of Arg52 with Gln54 may help position Gln54 for roles in two simultaneous pathways (Figure 6A, Step 3). On the one hand, stacking of Gln54 on G4 may position residues in the  $\beta$ 1-strand and  $\beta$ 1- $\alpha$ 1 loop for interaction with U2 and U3.



**Figure 6.** Tentative model for the sequential establishment of close-range interactions in the U1A/U1hpII complex. This working model is based on molecular dynamics simulations by us and others (46), structural data of the free and RNA-bound protein (9) and kinetic observations. The schematic on the left (A) indicates the hypothetical sequential progression of the interaction, in which the different steps are color-coded and numbered. The bases are indicated by circles, with the closing base pair marked at the top. Dashed lines indicate hydrogen bonds. Blue triangles designate water-mediated hydrogen bonds. Solid lines mark stacking interactions. Secondary structure features of the protein are indicated. The diagram on the right (B) shows the areas of the protein sequentially interacting with the RNA, coded in the same colors as (A). A section of the solution structure of the free protein (amino acids 2–101) (30) was used as the basis for the illustration in (B).

On the other hand, formation of a hydrogen bond from Gln54 to Tyr13 may position Tyr13 for stacking on C5. In turn, interactions of U3 and C5 with residues in the  $\beta$ 4-strand and adjacent loop may then induce the rearrangement of the C-terminal helix, which partially occludes the RRM surface (Figure 6A, Step 4), freeing Phe56 in  $\beta$ -strand 3 and Asp92 in helix C to stack on bases A6 and C7, respectively (Figure 6A, Step 5).

The kinetic effects of the mutations of Tyr13, Gln54 and Phe56 support this model; association defects are observed when Gln54 is mutated and molecular dynamics simulations indicate that mutation of Gln54 results in a repositioning of Arg52 and, in the case of the Gln54Glu mutant, a probable increase of negative charge in the vicinity of Arg52 that may reduce the attraction of the incoming RNA. The Gln54Glu mutation would thereby interfere with one of the earliest steps in association, which explains its marked impact on association. Loss of the aromatic side chain of Tyr13 may slow the rearrangement of the  $\beta$ 4-strand and adjacent loop, hindering the establishment of the Phe56–A6–C7–Asp92 stacking interaction, which would lead to the pronounced loss in association rate that we observed. Stacking onto Phe56 would thus be one of the last steps in complex formation, occurring coincident with or after rearrangement of the C-terminal helix. Replacement of Phe56 by Ala slows association, but much less so than a Tyr13Ser mutation, supporting the idea that stacking of an RNA base on Phe56 is a later event

in complex formation. Lastly, replacement of Tyr13 and Phe56 by other aromatic residues has a negligible effect on association, as expected.

It should be noted that the NMR and crystal structures used as starting points for our simulations are snapshots that may not fully represent the natural complex. In addition, our studies were carried out with fragments of the U1A protein and U1snRNA, and it remains to be determined whether these fragments interact in the same way as the full-length molecules. Nevertheless, the proposed model provides a useful framework for examining protein/RNA docking. Much further work will be required to confirm the proposed sequence of events during the locking of U1hpII and U1A RRM1. An extensive mutational analysis of Arg52 in the protein, and of the closing base pair and loop base A1 in the RNA will be important to further examine what may be the first close-range interaction in association. It would also be relevant to explore the consequences of adding more negative charge to the RRM surface, in the neighborhood of key residues. Such mutations would be predicted to slow association through repulsion, but the strength of such effects might depend on the location of the added charge. Examination of the kinetic consequences of removing the C-terminal helix would likewise be important. Its removal has been reported to cause a loss in affinity (50), but the kinetic consequences have not been analyzed. As helix C occludes part of the RNA-binding surface, the association of such a mutant with RNA would

probably be facilitated. However, any beneficial effects would probably be counteracted by a loss of complex stability associated with lack of the helix, perhaps resulting in a net loss in affinity. Preliminary kinetic analysis (M.J. Law, P.P. Anglim, O.A. Arreola and I.A. Laird-Offringa, manuscript in preparation) indicates that these expectations are correct and supports a model in which helix C first has to move away and then clamps down on the RNA.

Because an induced fit of protein and RNA appears to be a common theme in RNA-protein interactions, understanding the mechanism of complexation of a very well characterized complex, such as that of U1A/U1hpII, will help form models for less well defined interactions. Examples of such interactions are those between single-stranded RNA and multi-RRM proteins, such as the binding of the two N-terminal RRM of the neuronal HuD protein to AU-rich RNA (14). Based on the work presented here and the Tang and Nilsson model (46), it is reasonable to assume that certain positively charged residues might play a key role in first mediating electrostatic attraction and then initiating close-range contacts. HuD residues Lys111 and Arg116, which make highly specific contacts to adjacent U residues, are the most likely candidates for playing such a role. In support of this idea, these residues lie in the main RNA-binding domain, RRM1 (34,51). Rearrangement of a helix C-terminal to an RRM, similar to the process in U1A, has yet to be frequently observed in RRM-RNA interactions. This may be because fragments used for structural studies tend to include few residues downstream of the C-terminal RRM. One protein in which a C-terminal helix has been observed to adjust during RNA binding is CstF64, a factor involved in polyadenylation (52). The helix in question lies perpendicularly over the RRM surface and unfolds to provide access to the RNA. In another example, recent studies of the full-length (3-RRM) HuD protein suggest a rearrangement of a flexible hinge between RRM2 and 3 during RNA binding (S. Kim, K.C. Huang and I.A. Laird-Offringa, manuscript in preparation). The dissection of the mechanism of complex multi-RRM protein binding to RNA will require intensive investigation. In the meantime, the study of compact, well characterized interactions, such as those between U1A and U1hpII, might provide useful hints to possible steps in the formation of stable RRM-RNA complexes.

## ACKNOWLEDGEMENTS

We thank members of the Laird-Offringa lab for helpful criticism, and are grateful to Raveendra Dayam, John Barnes and Roland Beckmann for their kind assistance. This work was supported by National Science Foundation Grant MCB-0131782 (to I.A.L.-O.). Funding to pay the Open Access publication charges for this article was provided by National Science Foundation Grant MCB-0131782 to I.A.L.-O.

*Conflict of interest statement.* None declared.

## REFERENCES

- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Nagai, K., Oubridge, C., Ito, N., Avis, J. and Evans, P. (1995) The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.*, **20**, 235–240.
- Varani, G. and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing. *Ann. Rev. Biophys. Biomol. Struct.*, **27**, 407–445.
- Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
- Draper, D.E. (1999) Themes in RNA-protein recognition. *J. Mol. Biol.*, **293**, 255–270.
- Johansson, C., Finger, L.D., Trantirek, L., Mueller, T.D., Kim, S., Laird-Offringa, I.A. and Feigon, J. (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *J. Mol. Biol.*, **337**, 799–816.
- Katsamba, P.S., Myszkowski, D.G. and Laird-Offringa, I.A. (2001) Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J. Biol. Chem.*, **276**, 21476–21481.
- Birney, E., Kumar, S. and Krainer, A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein. *Nature*, **398**, 579–585.
- Deo, R.C., Bonanno, J.B., Sonenberg, N. and Burley, S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, **98**, 835–845.
- Allain, F.H.-T. (2000) Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. *J. Mol. Biol.*, **303**, 227–241.
- Allain, F.H.T., Howe, P.A., Neuhäus, D. and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5772.
- Wang, X. and Tanaka Hall, T.M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nature Struct. Biol.*, **8**, 141–145.
- Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R. and Xu, R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.*, **13**, 1102–1115.
- Price, S.R., Evans, P.R. and Nagai, K. (1998) Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, **394**, 645–650.
- Stump, W.T. and Hall, K.B. (1995) Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA*, **1**, 55–63.
- Kranz, J.K., Lu, J. and Hall, K.B. (1996) Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. *Protein Sci.*, **5**, 1567–1583.
- Deardorff, J.A. and Sachs, A.B. (1997) Differential effects of aromatic and charged residue substitutions in the RNA binding domains of the yeast poly(A)-binding protein. *J. Mol. Biol.*, **269**, 67–81.
- Nolan, S.J., Shiels, J.C., Tuite, J.B., Cecere, K.L. and Baranger, A.M. (1999) Recognition of an essential adenine at a protein-RNA interface: comparison of the contributions of hydrogen bonds and a stacking interaction. *J. Am. Chem. Soc.*, **121**, 8951–8952.
- Kranz, J.K. and Hall, K.B. (1998) RNA binding mediates the local cooperativity between the beta-sheet and the C-terminal tail of the human U1A RBD1 protein. *J. Mol. Biol.*, **275**, 465–481.
- Shiels, J.C., Tuite, J.B., Nolan, S.J. and Baranger, A.M. (2002) Investigation of a conserved stacking interaction in target site recognition by the U1A protein. *Nucleic Acids Res.*, **30**, 550–558.
- Lu, J. and Hall, K.B. (1995) An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J. Mol. Biol.*, **247**, 739–752.
- Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J. and Mattaj, I.W. (1989) Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.*, **8**, 4163–4170.
- Lutz-Freyermuth, C., Query, C.C. and Keene, J.D. (1990) Quantitative determination that one of two potential RNA-binding domains of the

- A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl Acad. Sci. USA*, **87**, 6393–6397.
26. Tsai, D.E., Harper, D.S. and Keene, J.D. (1991) U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res.*, **19**, 4931–4936.
  27. Hall, K.B. and Stump, W.T. (1992) Interaction of N-terminal domain of U1A protein with an RNA stem-loop. *Nucleic Acids Res.*, **20**, 4283–4290.
  28. Hall, K.B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry*, **33**, 10076–10088.
  29. Boelens, W.C., Jansen, E.J., van Venrooij, W.J., Stripecke, R., Mattaj, J.W. and Gunderson, S.I. (1993) The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, **72**, 881–892.
  30. Avis, J.M., Allain, F.H., Howe, P.W., Varani, G., Nagai, K. and Neuhaus, D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.
  31. Rupert, P.B., Xiao, H. and Ferre-D'Amare, A.R. (2003) U1A RNA-binding domain at 1.8 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 1521–1524.
  32. Kranz, J.K. and Hall, K.B. (1999) RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface. *J. Mol. Biol.*, **285**, 215–231.
  33. Katsamba, P.S., Park, S. and Laird-Offringa, I.A. (2002) Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods*, **26**, 95–104.
  34. Park, S., Myszka, D.G., Yu, M., Littler, S.J. and Laird-Offringa, I.A. (2000) HuD RNA recognition motifs play distinct roles in the formation of a stable complex with AU-rich RNA [Erratum (2004) *Mol. Cell Biol.*, **24**, 6888.]. *Mol. Cell Biol.*, **20**, 4765–4772.
  35. Williams, D.J. and Hall, K.B. (1996) RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein. *J. Mol. Biol.*, **257**, 265–275.
  36. Myszka, D.G. and Morton, T.A. (1998) CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.*, **23**, 149–150.
  37. Myszka, D.G., Jonsen, M.D. and Graves, B.J. (1998) Equilibrium analysis of high affinity interactions using BIACORE. *Anal. Biochem.*, **265**, 326–330.
  38. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Radmer, R.J., Duan, Y., Pitera, J., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1999) AMBER 6, University of California, San Francisco.
  39. Shafmeister, C.E.A.F., Ross, W.S. and Romanovski, V. (1995) *LeaP* University of California, San Francisco.
  40. Ryckaert, J.P., Cicciotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, **23**, 327–341.
  41. Darden, T.A., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
  42. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
  43. Katsamba, P.S., Bayramyan, M., Haworth, I.S., Myszka, D.G. and Laird-Offringa, I.A. (2002) Complex role of the  $\beta$ 2– $\beta$ 3 loop in the interaction of U1A with U1 hairpin II RNA. *J. Biol. Chem.*, **277**, 33267–33274.
  44. Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
  45. Pitici, F., Beveridge, D.L. and Baranger, A.M. (2002) Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: do molecular substrates code for specificity? *Biopolymers*, **65**, 424–435.
  46. Tang, Y. and Nilsson, L. (1999) Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.*, **77**, 1284–1305.
  47. Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nature Struct. Biol.*, **7**, 834–837.
  48. Mittermaier, A., Varani, L., Muhandiram, D.R., Kay, L.E. and Varani, G. (1999) Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.*, **294**, 967–979.
  49. Reyes, C.M. and Kollman, P.A. (2000) Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.*, **297**, 1145–1158.
  50. Zeng, Q. and Hall, K.B. (1997) Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA*, **3**, 303–314.
  51. Chung, S., Jiang, L., Cheng, S. and Furneaux, H. (1996) Purification and properties of HuD, a neuronal RNA-binding protein. *J. Biol. Chem.*, **271**, 11518–11524.
  52. Perez-Canadillas, J.-M. and Varani, G. (2001) Recent advances in RNA-protein recognition. *Curr. Opin. Struct. Biol.*, **11**, 53–58.